

Stichwort, Stichwortliste und Eigennamen in *lexiko*: Einflüsse der Korpusbasiertheit und Hypermedialität auf die lexikografische Konzeption

Annette Klosa, Ulrich Schnörch und Sabine Schoolaert
IDS Mannheim

*Die Überschrift des Beitrags impliziert dessen Gliederung in zwei größere thematische Abschnitte: der erste, allgemeine widmet sich Überlegungen zu Stichwort und Stichwortliste, der zweite, speziellere erörtert die Behandlung von Eigennamen in *lexiko*.*

**lexiko* (www.lexiko.de) ist ein am Institut für Deutsche Sprache in Mannheim entstehendes Online-Wörterbuch zur deutschen Sprache. Die methodische Basis für die redaktionelle, lexikografische Erarbeitung von Wortartikeln ist das Prinzip der Korpusbasiertheit. Voraussetzung für dessen methodische Umsetzung ist, dass für jedes Stichwort (und seine Lesarten) Belege in ausreichender Anzahl und Qualität im *lexiko*-Korpus vorhanden sind. Um das zu gewährleisten wurde auch die Stichwortliste komplett neu erstellt, und zwar auf der Basis von Korpora des geschriebenen Deutsch seit 1946. Im ersten Teil des Beitrags werden grundsätzliche Gedanken zur Erarbeitung einer adäquaten Stichwortkonzeption im Rahmen eines Online-Wörterbuches dargelegt, Sonderfälle und Ausnahmen vorgestellt sowie die Vorgehensweise bei der korpusbasierten Erstellung der *lexiko*-Stichwortliste skizziert.*

*Ausgehend von der Definition von Eigennamen erörtert der zweite Teil des Beitrags die gängige lexikografische Behandlung von Eigennamen in allgemeinsprachigen Wörterbüchern und stellt Überlegungen dazu an, wie Eigennamen in Abgrenzung zu Gattungsbezeichnungen lemmatisiert werden sollten. Dabei stellt sich für ein Online-Wörterbuch wie *lexiko*, dessen Schwerpunkt der lexikografischen Beschreibung auf der Bedeutung und Verwendung von Stichwörtern liegt, die Frage, in welcher Form die lexikografische Behandlung von Eigennamen erfolgen soll. Außerdem thematisiert dieser Beitrag die Behandlung von Eigennamen in *lexiko* hinsichtlich ihrer Erfassung, Klassifizierung und Darstellung und erläutert unterschiedliche Angabetypen. Ein Ausblick auf Suchoptionen zu den Eigennamen schließt die Überlegungen ab.*

lexiko (2003ff.) ist ein am Institut für Deutsche Sprache in Mannheim entstehendes Online-Wörterbuch zur deutschen Sprache, das über OWID, das Wörterbuchportal des Instituts für Deutsche Sprache (IDS), zugänglich ist (vgl. OWID 2008ff.).

Aus lexikografischer Sicht waren für die Erarbeitung der Stichwortliste und des Stichwortansatzes von *lexiko* (für Details vgl. Schnörch 2005a) vor allem zwei Faktoren ausschlaggebend, nämlich die hypertextuelle Struktur der Online-Publikation und das Prinzip der Korpusbasiertheit. In dem nachfolgenden, allgemein gehaltenen Abriss geht es vor allem um einige Konsequenzen und Neuerungen grundsätzlicher Art, die sich aus diesen beiden Grundgegebenheiten ableiten lassen.¹

Anders als im Printwörterbuch spielt das Alphabet im Rahmen der elektronischen Publikationsform von *lexiko* als ordnende Instanz keine Rolle mehr. Für den Arbeitsprozess, insbesondere die Artikelproduktion und die Suchstrategien werden andere Strukturierungs- und Recherchetechniken genutzt. Trotz Hypermedialität und den daraus resultierenden Möglichkeiten, flexiblere, weil dynamische und modulare Arbeitsstrategien zu entwickeln, ist es sinnvoll, auch bei einem 'elektronischen' Wörterbuchprojekt die Erstellung einer weitgehend kompletten Stichwortliste an den Anfang zu stellen: Eine Stichwortliste ist die Basis bzw. das Gerüst. Konzeptionell ist sie an den Erfordernissen des Projekts ausgerichtet, und zwar quantitativ im Hinblick auf den intendierten Umfang sowie qualitativ im Hinblick auf die Auswahl von Bearbeitungswortschätzen und auf die funktionale Definition der Einheit 'Stichwort', die aus den projektspezifischen Erfordernissen abzuleiten ist.

Um die Möglichkeiten des Hypertextes zu nutzen, ist u. a. ein Verlinkungssystem nötig. Im Prinzip ist dies mit den herkömmlichen Wörterbuchverweisen vergleichbar. Doch erst eine

¹ Im Vordergrund stehen dabei die lexikografischen, nicht die (text-)technologischen Aspekte.

vollständig angelegte, in einer Datenbank abgelegte und verwaltete Stichwortliste ermöglicht es, von Wortartikel a (ggf. Lesart x) auf Wortartikel b (ggf. Lesart y) zu verweisen. Die Stichwortliste bildet in diesem Kontext also gewissermaßen ein übergeordnetes ‘Adressverzeichnis’.

Ein für die Erarbeitung von Wortartikeln maßgebender Grundsatz in *elexiko* ist das Prinzip der Korpusbasiertheit: ‘Es bedeutet, dass in *elexiko* keine einzige Angabe enthalten ist, die allein aus der individuellen Sprecherkompetenz der LexikografInnen abgeleitet wurde. Die Angaben in *elexiko* werden strikt aus dem [...] regelmäßig aktualisierten *elexiko*-Korpus gewonnen [...]’ (Haß 2005a: 7). Voraussetzung für die methodische Umsetzung des Prinzips der Korpusbasiertheit ist, dass für jedes Stichwort (und seine Lesarten) Belege in ausreichender Anzahl und Qualität im *elexiko*-Korpus vorhanden sind. Daher wurde auch die Stichwortliste auf der Basis von Korpora des geschriebenen Deutsch seit 1946 in zwei Teilschritten komplett neu erstellt.

Der erste Schritt vom Korpus zur Stichwortliste erfolgte computergestützt mit Hilfe eines Lemmatisierers (vgl. Lemmatisierer 1994). Sein Einsatz erbrachte ca. 320.000 automatisch generierte Grundformen, die sog. Stichwortkandidatenliste. Hierbei handelt es sich um potenzielle *elexiko*-Stichwörter mit einer Mindestfrequenz von 8.

Im zweiten Teilschritt wurden diese Stichwortkandidaten nach redaktionellen, lexikologisch-lexikografischen Gesichtspunkten überprüft. Im Idealfall handelte es sich tatsächlich um eine wörterbuchrelevante Grundform im Sinne eines *elexiko*-Stichworts. In zahlreichen Einzelfällen musste aber auch nachgebessert werden. Im Zuge dessen wurden beispielsweise auch kuriose ‘Fehllemmatisierungen’, etwa durch Analogien fälschlicherweise generierte Substantivierungen wie *Alphornblase*, *Augenreibe*, *Denkmalgeschütz* und *Drachenfliege* ausgemustert. Ergebnis dieser Sichtung ist die *elexiko*-Stichwortliste mit gut 300.000 Einheiten.

Voraussetzung für diese ebenso zeitaufwendige wie notwendige Prüfung sind natürlich Lemmatisierungsprinzipien, sprich Leitlinien, die den Lexikografen und Lexikografinnen Aufschluss darüber geben, welche Wörter mit welcher (Grund-)Form als Stichwort in *elexiko* anzusetzen und welche Varianten zu berücksichtigen sind. Dabei ist – wie bereits angedeutet – zu beachten, dass die elektronische Publikationsform von *elexiko* andere Anforderungen an das Konzept des Stichwortes stellt als jene, die für das ‘Lemma’ der traditionellen Printlexikografie formuliert wurden (vgl. u. a. Wiegand 1983). In der Makrostruktur eines gedruckten Wörterbuches dienen die alphabetisch angeordneten Stichwörter v. a. der Orientierung bzw. der Suche. Nutzer und Nutzerinnen finden Informationen entweder direkt bei einem Stichwort in seiner ‘kanonischen’ Grundform, oder sie werden von ausgewählten lemmatisierten Flexionsformen auf die lemmatisierte Grundform verwiesen, z. B. von *sah* zu *sehen*.

Dieses Suchen und Verweisen erfolgt im hypertextuellen Rahmen von *elexiko* natürlich über ein (im Aufbau befindliches) Recherchesystem bzw. über Links. Unter diesem Aspekt reduziert sich die Funktion des Stichwortes in *elexiko* in erster Linie auf die der möglichst eindeutigen Etikettierung. Das Lemmazeichen bildet als Lemmazeichengestaltungsgabe gewissermaßen die Überschrift für die XML-Instanz des Wortartikels. Unregelmäßige (einem bloßen Verweis dienende) Flexionsformen sucht man daher meist vergeblich in der *elexiko*-Stichwortliste.

Verallgemeinert kann folgendes Prinzip für die Stichwortansetzung in *ellexiko* formuliert werden: Oberstes Kriterium für den Status eines Stichwortes bei *ellexiko* ist seine ‘Eigenständigkeit’: Neben den lemmatisierten ‘Grundformen’ i. e. S. findet man nur solche Wortformen, die sich inhaltlich und/oder formal nicht über diese Grundform (bzw. den dazugehörigen Wortartikel) ‘erschließen’ und beschreiben lassen, bzw. für die eine Grundform im Korpus nicht (oder nur unzureichend) belegt ist. Neben diesem funktionalen Unterschied erfolgt der Stichwortansatz der Hauptwortarten in formaler Hinsicht natürlich in gleicher Weise wie in Printwörterbüchern, d.h. Nomina werden im Nominativ Singular, Adjektive im Positiv, und Verben im Infinitiv lemmatisiert usw.²

Ausnahmen, die im Rahmen des ursprünglichen Stichwortkonzepts eigens behandelt werden müssen, sind u. a. Pluraliatantumformen (wie *Stadtväter* oder *Flitterwochen*) und bestimmte Formen des Partizips 1 und 2 (z. B. *ätzend* oder *begabt*, vgl. dazu auch Erb 2005). Für drei weitere Gruppen wurden erst jüngst Ausnahmeregelungen festgeschrieben, dabei handelt es sich um bestimmte Adjektivgruppen im Bereich der Autosemantika sowie aus dem Bereich der Synsemantika um Pronomen und Artikel(wörter).

Adjektive und lemmatisierte Steigerungsformen, für die es keine verbindliche Grundform gibt, werden unter einer genusunspezifischen, endungslosen Form mit einem Platzhalterbindestrich zusammengefasst, also etwa *äußerst-* (für *äußerste*, *äußerster*, *äußerstes*) bzw. *best-* (für *beste*, *bester*, *bestes*). Das allgemeine Lemmatisierungskonzept sah zwar zunächst vor, auf Platzhalter bei der Lemmazeichengestaltangabe zu verzichten (zumal Druckraumknappheit in einem Online-Wörterbuch praktisch nicht existiert und die Explizitheit nicht beschränkt). Nach reiflicher Überlegung wurde in den genannten Fällen jedoch eine Entscheidung gegen diese Vorgabe getroffen. Drei getrennte Stichwörter *äußerste*, *äußerster*, *äußerstes* anstelle von *äußerst-* anzusetzen hieße in der Konsequenz, auch drei *ellexiko*-Wortartikel vollständig redaktionell zu erarbeiten, (genusspezifische) Belege zu recherchieren, für eine adäquate Vernetzung zu sorgen usw. Der Arbeitsaufwand wäre erheblich, nur um letztlich ein Maß an lexikografischer Redundanz zu produzieren, das in keinem Verhältnis zum tatsächlichen qualitativen Mehrwert an Information stünde. Mit der gewählten Darstellungsform wird demnach auf sinnvolle Weise der Tatsache Rechnung getragen, dass die Formen aller Genera sich in ihrer Bedeutung und Verwendung nicht unterscheiden. Bei dieser Gruppe von Autosemantika liegt der Fokus im Wortartikel auf den semantischen Angaben.

Bei den Artikelwörtern und Pronomen hingegen liegt der Schwerpunkt auf ihrer funktionalen und insbesondere ihrer grammatischen Beschreibung, sodass hier eine getrennte Lemmatisierung vorzuziehen ist: Bei den Artikeln und Pronomen werden also jeweils alle Formen des Nominativ Singular als einzelne Lemmata angesetzt (z.B. *der*, *die*, *das*; *dein*, *deiner*, *deine*, *deines*). Der Zusammenhang dieser Artikel wird über entsprechende Formulierungen (in der semantischen Paraphrase wird auf die Formen in den anderen Genera hingewiesen) sowie über Hyperlinks gewährleistet.³

Für das Stichwortkonzept des Online-Wörterbuches *ellexiko* war von Anfang an folgender Leitgedanke maßgebend: Die Lemmazeichengestaltangabe dient v. a. der eindeutigen Etikettierung eines Wortartikels, in dem alle korpusbasiert ermittelten relevanten

² Wortbildungselemente, Mehrworteinheiten, Siglen, die meisten Abkürzungen usw. blieben bis dato unberücksichtigt! Zum Teil fällt dies in den Aufgabenbereich anderer Projekte im Rahmen des Portals OWID.

³ Zusätzlich wird auf den entsprechenden Eintrag in ‘grammis’, dem grammatischen Informationssystem des IDS, verlinkt, wo weitere morphologische, syntaktische und semantische Informationen aufgerufen werden können.

Informationen unter verschiedenen Schwerpunkten zusammengefasst sind. In *lexiko* existieren folglich keine zwei völlig identischen Lemmazeichengestaltungangaben.⁴ Auch die im Projektfortgang notwendig gewordenen Konkretisierungen und Ausnahmeregelungen, von denen einige hier vorgestellt wurden, bewegen sich innerhalb dieses Rahmens.

Nachdem nun skizziert wurde, wie sich Hypermedialität und Korpusbasiertheit allgemein auf das Stichwortkonzept und auf die Stichwortliste eines Online-Wörterbuches auswirken können, soll in den folgenden Abschnitten erörtert werden, wie *lexiko* mit einem ganz konkreten, für Wörterbücher gemeinhin nicht ganz unproblematisch zu handhabenden Wortschatzausschnitt in puncto Stichwortliste und Stichwortansatz verfährt, nämlich den Eigennamen.

Eigennamen werden neben Gattungsnamen und Stoffnamen als Unterart des Nomens klassifiziert (Zifonun et. al. 1997: 32) und leisten Folgendes:

Eigennamen dienen der konstanten Bezeichnung bestimmter Individuen (insbesondere von Personen, Orten, Ländern, Regionen, Flüssen, Waren usw.) gemäß einer (mindestens zu unterstellenden) Vereinbarung (Taufakt usw.). Besteht eine entsprechende Verbindung zwischen Individuum und Eigenname und kann sie im Sprecher- und Hörerwissen abgerufen werden, so ist mit der Verwendung des Eigennamens ein Verweis auf den Träger unproblematisch möglich. Eigennamen haben – anders als Gattungs- und Stoffnamen – keine Charakterisierungsfunktion, sie erlauben allenfalls Rückschlüsse auf wenige Eigenschaften der Individuen (etwa das Geschlecht). (Zifonun et. al. 1997: 32)

Das heißt, Eigennamen werden in der Regel als situationsunabhängiges Referenzmittel definiert, das sich auf ein Individuum bezieht und über den bloßen referentiellen Charakter hinaus keine lexikalische Bedeutung hat.

Die lexikografische Behandlung von Eigennamen erfolgt üblicherweise in eigenen Nachschlagewerken (Namenbüchern, Namenlexika, Namenatlanten). Dennoch hat sich die Tradition herausgebildet, vor allem Personennamen und geografische Namen in allgemeinsprachliche Wörterbücher aufzunehmen. Dies liegt zum einen daran, dass es ein Nachschlagebedürfnis nach Eigennamen gibt, z. B. in Bezug auf ihre Schreibung. Zum anderen haben

Eigennamen einen nicht unerheblichen Anteil am deutschen Wortschatz, den man nicht einfach ignorieren kann.

Für *lexiko*, ein Online-Wörterbuch, das streng korpusbasiert arbeitet und dessen Schwerpunkt auf der Bedeutung und Verwendung von Stichwörtern liegt, ergeben sich hieraus folgende Überlegungen: Sollen Eigennamen nur dann lemmatisiert werden, wenn sie auch als Gattungsbezeichnungen verwendet werden? In welcher Form sollen Eigennamen dargestellt werden und welche Informationen sollen Nutzer zu Eigennamen erhalten bzw. welche Informationen sind sinnvoll?

Werden Eigennamen in Wörterbücher aufgenommen, dann ist es wichtig, dass sich deren Darstellung und Beschreibung von der anderer Stichwörter unterscheidet, auch im Hinblick auf die Bedeutungserläuterung:

⁴ Auch die Trennung zwischen Polysemie und Homonymie ist in *lexiko* grundsätzlich aufgehoben, es gibt also keine homonymen Stichwortansätze. Stattdessen werden Zusammenhänge zwischen Lesarten eines Stichwortes unter der Rubrik 'Zum Zusammenhang der Lesarten' beschrieben. Begünstigt, wenn nicht gar ermöglicht wird dieses Vorgehen freilich erst durch die hypertextuelle Präsentation der Wortartikel.

Section 3. Reports on Lexicographical and Lexicological Projects

[...] nimmt man die Eigennamenforschung ernst [...], dann muß in der lexikographischen Praxis – soll der Benutzer nicht irregeführt werden – das nomen proprium grundsätzlich anders beschrieben werden als das nomen appellativum. Dies gilt nicht nur für die morphologischen Angaben, sondern vor allem für die Bedeutungserläuterungen, die bei Eigennamen einen ganz anderen Status und eine andere Funktion haben als bei Gattungsbezeichnungen [...]. (Strauß 1989: 788)

Diese Trennung wird jedoch nicht immer eingehalten, wie Harweg (1999) anhand verschiedener Beispiele im ‘Duden – Die Deutsche Rechtschreibung’ zeigt. Harweg unterscheidet zwischen metasprachlichen (das Sprachwissen betreffenden) und nichtmetasprachlichen (das Sachwissen betreffenden) Einträgen und macht dies an verschiedenen Beispielen deutlich. So werden im Rechtschreibduden einige biblische Vornamen mit metasprachlichen Einträgen versehen, wie z. B. ‘*Abraham, biblischer männlicher Eigename*’, andere wiederum mit nichtmetasprachlichen Einträgen, wie z. B. ‘*Moses, jüdischer Gesetzgeber im Alten Testament*’. Auch die Einträge zu Familiennamen weisen Inkonsistenzen auf. So sind die Einträge zu *Böcklin (schweizerischer Maler)* und *Heidegger (deutscher Philosoph)* in nichtmetasprachlicher Form verfasst, während sie zu *Humboldt (Familiennamen)* in metasprachlicher Form formuliert sind (Harweg 1999).

Es gibt verschiedene Gründe, die für eine lexikografische Behandlung von Eigennamen sprechen. Praktische Gründe sind das Nachschlagebedürfnis der Nutzer in Bezug auf Orthografie und Grammatik (z. B. Unsicherheiten bei Genus und Numerus). Doch auch der interessante Aspekt der Verschränkung zwischen Gattungsbezeichnungen und Eigennamen und die Schwierigkeit, sie voneinander abzugrenzen, sprechen dafür, Eigennamen lexikografisch zu erfassen:

Für die Korpuslinguistik ist die verlässliche Unterscheidung zwischen Appellativa und Propria ein Desiderat, das *ellexiko* schließen helfen kann. Und für die lexikalische Semantik ergeben sich vielleicht neue Fragen in Bezug auf die Verschränkung von Appellativa und Propria, die insbesondere bei Gruppen-, Firmen- und Institutionennamen keineswegs nur ausdrucksseitig (z.B. wie bei *Mark, Kohl* und *Schneider*), sondern semantisch-pragmatisch relevant ist. (Haß 2005b: 168)

Die Eigennamen in *ellexiko* sind durchweg Einwortlemmata; mehrteilige Namen werden grundsätzlich nicht lemmatisiert (z. B. *Karl der Kahle, Französische Revolution*), außerdem keine Abkürzungen, die als Name fungieren (z. B. *USA*). Da Eigennamen auf ein einziges Objekt in der Welt Bezug nehmen (d. h. referieren), gehören sie in *ellexiko*, zusammen mit anderen referierenden Wörtern wie z. B. den Pronomina, in die Wortklasse der Referenzwörter. Die Klassifikation in eine der Wortklassen bedingt in *ellexiko*, wie die Paraphrase zu formulieren ist, welche syntagmatischen und paradigmatischen Beziehungen zu erwarten sind und welche grammatischen Angaben gemacht werden. In der *ellexiko*-Artikelstruktur wird der jeweilige Namentyp erfasst, was vor allem für differenzierte Rechercheoptionen wichtig ist.

Wie eingangs schon erwähnt wurde, ist *ellexiko* ein Wörterbuch, dessen Erarbeitung dem Prinzip der Korpusbasiertheit folgt. Die wichtigste Entscheidung bezüglich Eigennamen in *ellexiko* lag folglich auch darin, dass sie nicht aus der korpusbasiert komplett neu erstellten Stichwortliste aussortiert wurden. Begründet wurde dies hauptsächlich mit ihrer Vorkommenshäufigkeit (Stichwortkandidaten mit einer Korpusfrequenz unter 8 fanden allerdings generell keinen Eingang in die Stichwortliste). Bei der Erstellung der *ellexiko*-Stichwortliste wurden 3.763 Stichwortkandidaten als Namen markiert, von denen nach redaktioneller Prüfung 3.689 Stichwörter wirklich als Namen klassifiziert wurden, was ungefähr einem Prozent der Stichwörter in *ellexiko* entspricht. Diese Namen verteilen sich folgendermaßen auf die einzelnen Namentypen:

Namentyp	Anzahl
Familiennamen	1334
Ortsnamen	1065
Namen von Regionen und Landschaften	332
Vornamen/Rufnamen/historische Individualnamen	267
Staatennamen	189
Namen von Gebäuden	176
mythologische/literarische/religiöse Namen	171
Namen von Inseln	106
Namen von Flüssen	87
Namen von Bergen	85
Flurnamen	78
Firmennamen	70
Produkt- und Warennamen	48
Namen von Texten, Filmen, Musik, Gemälden	38
Namen von Himmelskörpern	35
Namentyp	Anzahl
Künstlernamen	30
Namen von Institutionen/Vereinigungen/Organisationen	29
Namen von Fortbewegungsmitteln	24
Namen von Seen	21
Namen von Gebirgen	19
Namen von Straßen und Plätzen	17
Namen von Kontinenten	10
Namen von Meeren	8
Völker- und Stammesnamen	8
Namen von Ereignissen und militärischen Aktionen	6
Namen von Naturphänomenen	2

Tab. 1. Häufigkeit der Namentypen in der gesamten *elexiko*-Stichwortliste

In einem weiteren Schritt wurde der zur Zeit redaktionell lexikografisch in *elexiko* bearbeitete Teilwortschatz auf Eigennamen hin untersucht. Dabei handelt es sich um das ca. 2700 hochfrequenten Stichwörter umfassende ‘Lexikon zum öffentlichen Sprachgebrauch’ (vgl. auch http://www.owid.de/elexiko_/modSpachgeb.html). Diese Überprüfung ergab, dass es rund 100 Eigennamen enthält, z. B. *Europa, Deutschland, Stuttgart, Hans, Hitler, Siemens*, aber auch Familiennamen wie *Weber, Töpfer, Fischer, Stich*, die ebenfalls als Gattungsbezeichnung vorkommen.

Daneben wurden Eigennamen im ‘Lexikon zum öffentlichen Sprachgebrauch’ im Zusammenhang mit der Bearbeitung von Gattungsbezeichnungen sichtbar, wenn die Durchsicht der Liste an Kookkurrenzpartnern zu dem jeweiligen Stichwort im Zuge der Lesartendisambiguierung (vgl. Storjohann 2004, Schnörch 2005b: 112ff. und Schnörch 2010) entsprechende Hinweise lieferte (z.B. *Sims*: Familienname, *Garant*: Name einer Firma, *August*: männlicher Vorname und Familienname).

Zu Beginn der Erarbeitung des ‘Lexikons zum öffentlichen Sprachgebrauch’ wurde versucht, die Eigennamen im Rahmen der auch für Gattungsbezeichnungen vorgesehenen Artikelstruktur zu erfassen und zu beschreiben. Dabei traten Probleme auf, die verschiedene Bereiche wie Online-Präsentation, Terminologie oder auch das methodische Vorgehen betrafen:

Die wichtigste Methode bei der lexikografischen Arbeit in *elexiko* ist die Korpusbasiertheit, d. h. alle Angaben in *elexiko* werden aus dem zugrundeliegenden Korpus gewonnen. Dabei kommen zwei Vorgehensweisen zum Einsatz: die korpusgeleitete sowie die korpusbasierte

Ermittlung von Angaben. Bei dem korpusgeleiteten Ansatz kommt vor allem das Verfahren der statistischen Kollokationsanalyse (vgl. Belica 1995) zum Tragen, das in COSMAS II implementiert ist. Die Kollokationsanalyse liefert dem Lexikografen erste statistisch signifikante Informationen zu einem Stichwort, wie z.B. die syntagmatischen oder paradigmatischen Beziehungen, die dann in einem weiteren Schritt entsprechend ausgewertet werden. In Bezug auf Eigennamen ist dieses Verfahren jedoch wenig hilfreich, da die durch die Kollokationsanalyse ermittelten Kookkurrenzpartner nur statistisch Relevantes zum Namensträger bzw. zum Denotat, nicht aber zum Namenwort selbst aussagen.

Bei der Entscheidung, generell die gleichen Online-Ansichten für alle Wortklassen vorzusehen, wurde insbesondere bei den Eigennamen in Kauf genommen, dass bestimmte Angabebereiche (z. B. ‘Semantische Umgebung’, ‘Typische Verwendungen’) unausgefüllt bleiben, was nicht besonders nutzerfreundlich ist. Außerdem wurde die für die Oberfläche gewählte Terminologie den Eigennamen nicht gerecht. So ist z. B. der Terminus ‘Lesart’ für Eigennamen nicht zutreffend, da ein Lemmazeichen, das auch als Name gebraucht wird, nicht eine Lesart bzw. Einzelbedeutung dieses Lemmazeichens ist. Auch der Terminus ‘Bedeutungserläuterung’ ist in Bezug auf Eigennamen kritisch, geht man davon aus, dass Eigennamen keine Bedeutung haben. Hinzu kommt die Frage, welchen Informationswert Paraphrasen wie ‘Müller ist ein Familienname’ oder ‘Bulgarien ist der Name eines Staates in Europa’ haben.

Vor dem Hintergrund dieser praktischen Erfahrungen ist deshalb vorgesehen, Eigennamen in *lexiko* grundsätzlich anders zu behandeln als Gattungsbezeichnungen, wobei hier zwischen Namen, die nicht zugleich als Gattungsprädikator belegt sind, und solchen, bei denen das Lemmazeichen auch als Gattungsprädikator belegt ist, unterschieden werden muss. Eigennamen, die in den Belegen des *lexiko*-Korpus nicht zugleich als Gattungsprädikator vorkommen, werden nur mit automatisch generierten Angaben (insbesondere mit orthografischen Angaben und Belegen) versehen (vgl. Abb. 1, die die tatsächliche Online-Bildschirmansicht des Wortartikels *Stuttgart* in für den Druck angepasster Form wiedergibt).

Stuttgart

Orthografie

Normgerechte Schreibung: Stuttgart
Worttrennung: Stutt|gart

Belege (automatisch ausgewählt)

Bahnsprecher Martin Schmolke teilt mit, die Planungen der Bahn liefen darauf hinaus, die Interregio-Linie Saarbrücken - Mannheim beizubehalten, die Verbindung nach **Stuttgart** aber ‘wegen der parallel laufenden ICE- und IC-Züge’ zu streichen. [...] Derzeit fahren IR-Züge von Mannheim ab 7.11 Uhr alle zwei Stunden nach **Stuttgart**. Sechs Interregio-Verbindungen sind es täglich, ebenso viele in der Gegenrichtung. (M98/DEZ.95921 Mannheimer Morgen, 22.12.1998, Ressort: Lokal Mannheim; Nach Stuttgart nicht mehr im Interregio)

[...] unterschiedliche Löhne gibt es überdies in der Bundesrepublik nicht nur zwischen den verschiedenen Branchen, sondern auch zwischen den Regionen. denn ein Metallarbeiter in **Stuttgart** hat schon immer mehr Lohn kassiert als sein Kollege in Leer, weil der Unternehmer im Südwesten den Tariflohn stärker aufstocken kann als der Firmenchef in Ostfriesland. dies scheint, wie vieles andere, den beiden Liberalen nicht gegenwärtig gewesen zu sein. (H85/BZ1.16526 Die Zeit, 19.04.1985, S. 23; An der Sache vorbei)

[...]

Dieses Stichwort gehört im *lexiko*-Korpus der Frequenzschicht X (50.001-100.000 mal belegt) an. Es ist in 17 verschiedenen Zeitungen oder Zeitschriften aus 24 Jahrgängen belegt.

Abb. 1: Wortartikel *Stuttgart* mit automatisch generierten Angaben

Bei den Stichwörtern, die sowohl als Eigenname wie als Gattungsbezeichnung im Korpus belegt sind, werden nur die Lesarten für die Gattungsprädikatoren lexikografisch erarbeitet, z.B. bei dem Familiennamen *Müller* nur die Lesart 'Berufsbezeichnung' oder bei den Staatennamen wie z. B. *Brandenburg* gegebenenfalls eine Lesart 'Gesamtheit der Personen'. Um jedoch die Erwartungshaltung der Nutzer nicht zu enttäuschen, die bei einem Wort wie *Brandenburg* sicherlich nicht in erster Linie die Lesart 'Gesamtheit der Personen' erwarten wird unter der Überschrift 'Zu den Lesarten' vermerkt, dass es sich bei *Brandenburg* auch um den Namen eines Bundeslandes oder eines Ortes handelt.

In der Übersichtsliste zu den Lesarten des Stichwortes erscheinen die Namen außerdem mit einer Etikettierung. Diese Etikettierung enthält genau jene Informationen, die der Nutzer dazu benötigt, festzustellen, um welche Art von Eigennamen es sich handelt (z. B. 'Name eines Ortes', 'Name eines Produkts', 'Familiennamen'; vgl. Storjohann 2005: 199ff.). Auf die Paraphrase und die Bezeichnung 'Lesart' in Bezug auf Eigennamen wird verzichtet. Es lassen sich auch keine weiteren Informationen zu den Namen abrufen. Der Nutzer erfährt also nur, dass es sich bei dem Stichwort auch um einen Eigennamen handelt und welche Art von Eigenname vorliegt.

Liegt eine Verschränkung zwischen Gattungs- und Eigenname vor, so wird zukünftig unter der Überschrift 'Zu den Lesarten' darauf hingewiesen. Dies betrifft sowohl Gattungsbezeichnungen, die sich aus Namen entwickelt haben (z. B. aus dem Namen *Zeppelin* die allgemeine Bezeichnung für ein Luftschiff) als auch Namen, die sich aus Gattungsbezeichnungen entwickelt haben. Dies ist häufig bei Familiennamen aus Berufsbezeichnungen (z. B. *Fischer*, *Weber*), bei Familiennamen als Übernamen zu Adjektiven (z. B. *Grün*, *Schön*) sowie bei Firmennamen (z. B. *Garant*, vgl. Abb. 2) der Fall.

<p>Garant Lesart 'Gewähr' Name einer Firma</p> <p>Zu den Lesarten Zu der Lesart 'Gewähr' ist der im elexiko-Korpus belegte Name der Firma Garant entstanden (vgl. den Beleg).</p> <p>Die drei Großbetriebe Garant (Mischfutter), Agrana (Stärke) und VOG (Vereinigung österreichischer Großzuckerhändler) galten als Verursacher von Gestank in Aschach und Hartkirchen. Agrana und Garant sanierten ihre Anlagen, auch die VOG ließ sich für ihre Rapsöl-Produktion eine Abluft-Nachverbrennung 1996 bewilligen. (Oberösterreichische Nachrichten, 10.11.1999, Umweltschutz hat Warten auf Taten der Aschacher VOG satt.)</p>

Abb. 2. Wortartikel *Garant* mit Informationen zum Eigennamen

Außerdem können Belege erscheinen, die den Namen in einem typischen Kontext zeigen, der möglichst den Namentyp verdeutlicht und darüber hinaus enzyklopädische Informationen liefert. Enzyklopädische Informationen zum Namensträger werden auch unter der Überschrift 'Weitere Informationen' aus sachbezogenen Nachschlagewerken zitiert (vgl. den Wortartikel *Brandenburg* in Abb. 3).

Brandenburg

[Lesart 'Gesamtheit der Personen'](#)

Name eines deutschen Bundeslandes

Name eines Ortes

Zu den Lesarten

Brandenburg ist als Name eines deutschen Bundeslandes im *lexiko*-Korpus belegt. Die Lesart 'Gesamtheit der Personen' ist zu diesem Namen eine Metonymisierung.

Weitere Informationen:

'Brandenburg, Land im Osten der Bundesrepublik Deutschland, mit 29476 km² flächengrößtes der neuen Bundesländer, (1999) 2,59 Mio. Einwohner, Hauptstadt ist Potsdam. Brandenburg grenzt im Osten mit Lausitzer Neißer und Oder an Polen. Inmitten von Brandenburg liegt das Land Berlin.' (Brockhaus in Text und Bild Edition, Bibliographisches Institut & F.A. Brockhaus AG, 2002 (CD-ROM)).

Brandenburg ist als Name eines Ortes im *lexiko*-Korpus belegt. Die Lesart 'Gesamtheit der Personen' ist auch zu diesem Namen eine Metonymisierung.

Weitere Informationen:

'Brandenburg an der Havel, kreisfreie Stadt in Brandenburg, beiderseits der Havel, 79000 Einwohner; Fachhochschule, OLG, Theater, Museen; Elektrostahlwerk, Druckmaschinenbau, Bauwirtschaft und Nahrungsmittelindustrie. Verkehrsknotenpunkt mit Hafen am Elbe-Havel-Kanal.' (Brockhaus in Text und Bild Edition, Bibliographisches Institut & F.A. Brockhaus AG, 2002 (CD-ROM)).

Abb. 3. Wortartikel *Brandenburg* mit lexikografisch erarbeiteten Informationen zu den Eigennamen

Ausblick

Der im ersten Teil vermittelte Einblick in die Erarbeitung der Stichwortliste sowie die Informationen zur Art und Genese des *lexiko*-Stichwortkonzeptes haben gezeigt, dass die Online-Publikation in Verbindung mit dem Prinzip der Korpusbasiertheit ein höheres Maß an Flexibilität und Dynamik bedingt und ermöglicht. Aufgrund der Erfahrungen im Laufe der praktischen Artikelarbeit in den vergangenen Jahren lässt sich zu folgendem Schluss kommen: Die Stichwortliste wird sich auch weiterhin verändern, es wird einzelne Ergänzungen und Streichungen geben, aber auch über die systematische Aufnahme ganzer Typen bestimmter Stichwörter, z.B. Abkürzungen, wird in absehbarer Zeit nachzudenken sein.

In Zukunft sollen in *lexiko* des Weiteren spezifische Abfragen sowohl über die lexikografisch bearbeiteten als auch die noch unbearbeiteten Stichwörter hinsichtlich der Eigennamen entwickelt werden: So könnte die Klassifikation der Namen Grundlage für eine Recherche über die noch unbearbeiteten Stichwörter nach einzelnen Namentypen und außerdem nach nur einfach und mehrfach klassifizierten Eigennamen ermöglichen. Solche Recherchen könnten insofern interessant sein, als sie zum ersten Mal einen Eindruck von der Verteilung bestimmter Namentypen in großen, zeitungssprachlichen Korpora vermitteln können.

Auch der Wunsch, das Phänomen der Verschränkung zwischen appellativem und proprietärem Wortschatz sichtbar zu machen, könnte in verschiedenen Bereichen umgesetzt werden. Vorgesehen ist in *lexiko* bisher, dass der Nutzer in der Online-Ansicht der redaktionell bearbeiteten Artikel unter dem Angabebereich 'Zu den Lesarten' Informationen darüber erhält, ob eine Verschränkung zwischen Eigenname und Gattungsbezeichnung vorliegt. Zukünftig denkbar ist, dass eine Suche über die redaktionell bearbeiteten Stichwörter nach solchen Eigennamen, zu denen durch Metonymisierung Gattungsbezeichnungen entstanden sind, oder auch nach Fällen, in denen ein Eigenname zu einer

Gattungsbezeichnung entstanden ist, sicherlich interessante, korpusgestützte Beispiele liefern könnte.

Literatur

Forschungsliteratur

- Erb, S. (2005). 'Stichwortansetzung von Partizipien'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 91-95.
- Harweg, R. (1999). *Studien zu Eigennamen: Eigennamen als Einträge in Wörterbüchern und Lexika.* Aachen: Shaker Verlag. 61-123.
- Haß, U. (2005a). 'elexiko – Das Projekt'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 1-17.
- Haß, U. (2005b). 'Das Bedeutungsspektrum'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 163-181.
- Schnörch, U. (2005a). 'Die Stichwortliste'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 71-90.
- Schnörch, U. (2005b). 'Der Autoren-Arbeitsplatz: Ein elexiko-Wortartikel entsteht'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 105-130.
- Storjohann, P. (2004). 'Computergestützte Lesartendisambiguierung'. In *Deutsche Sprache* 1/03. 3-28.
- Storjohann, P. (2005). 'Semantische Paraphrasen und Kurzetikettierungen'. In Haß, U. (Hg.). *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz.* (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York: de Gruyter. 182-203.
- Strauß, G. (1989). 'Angabe traditioneller Wortarten oder Beschreibung nach funktionalen Wortklassen im allgemeinen einsprachigen Wörterbuch?'. In Hausmann, F. J. (Hg.). *Wörterbücher: Ein internationales Handbuch zur Lexikographie.* 1. Teilband. Berlin: de Gruyter. 788-796. (HSK Bd. 5.1)
- Wiegand, H.-E. (1983). 'Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung'. In Wiegand, H.-E. (Hg.). *Studien zur neuhochdeutschen Lexikographie III.* (= Germanistische Linguistik 1-4/82). Hildesheim/Zürich/New York: Georg Olms. 401-474.
- Zifonun, G.; Hoffmann, L.; Strecker, B. (1997). *Grammatik der deutschen Sprache.* Band 3. (= Schriften des IDS Bd. 7.3). Berlin/New York: de Gruyter.

Wörterbücher

Duden – *Die deutsche Rechtschreibung.* 25., völlig neu bearbeitete und erweiterte Auflage. Mannheim u. a.: Duden, 2009.

Internetquellen

- Belica, C. (1995). Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analyseverfahren. <http://corpora.ids-mannheim.de/ccdb> (zuletzt besucht am 25.02.2010).
- elexiko (2003ff.). <http://www.elexiko.de> (zuletzt besucht am 23.02.2010).
- grammis – das grammatische Informationssystem des Instituts für Deutsche Sprache. <http://hypermedia.ids-mannheim.de/index.html> (zuletzt besucht am 23.02.2010).
- Lemmatisierer (1994). <http://www.ids-mannheim.de/kt/dokumente/glemmrep.pdf> (zuletzt besucht am 23.02.2010).
- OWID – Online-Wortschatz-Informationssystem Deutsch (2008ff.). Hrsg. v. Institut für Deutsche Sprache, Mannheim. <http://www.owid.de> (zuletzt besucht am 23.02.2010).
- Schnörch, U. (2010). 'Lesartendisambiguierung in elexiko'. <http://dwb.bbaw.de/tagung09/> (zuletzt besucht am 14.01.2010).